## CLAIMS

Having thus described our invention, what we claim as new and desire to secure by Letters Patent is as follows:

1. A method of learning annotators for use in an interactive machine learning system, the method comprising the steps of:

providing at least partially annotated text data or unannotated text data with seeds or seed models of instances of at least one named entity or class to be learned;

iteratively learning annotators for the at least one named entity or class using a machine learning algorithm;

applying the learned annotators to text data resulting in the annotation of at least one named entity or class annotation instance; and

selectively presenting for review and correction, if determined, representations of the at least one named entity or class annotation instance identified by the applying of the learned annotators.

2. The method of claim 1, wherein the annotations instances are selectively presented for review and correction, if determined, based on a predetermined threshold value of a confidence level.

3. The method of claim 1, wherein the step of iteratively learning includes incrementally improving the learned annotators.

4. The method of claim 1, wherein the at least one named entity is any syntactic, semantic or notional type that can be identified as a type and named.

5. The method of claim 1, wherein the seeds or seed models are at least one of lists, dictionaries, glossaries, patterns and database entries.

6. The method of claim 1, further comprising providing a log of corrections of removed or altered annotation instances.

7. The method of claim 6, wherein the log of corrections are optionally used to override any of the at least one named entity or class annotation instance inconsistent with the log.

8. The method of claim 1, further including preprocessing groups of words or phrases into single units before the iteratively learning step.

9. The method of claim 1, wherein the applying step provides confidence levels for each annotation instance such that the learned annotators and their respective confidence levels are used to selectively present some of the representations of the at least one named entity or class annotation instance.

10. The method of claim 9, wherein if confidence levels do not fall within a closed interval then a transformation will be applied to map a confidence level range onto the closed interval [0 ... 1 ] for purposes of presentation to the user.

11. The method of claim 9, further including adjusting a threshold of the confidence levels associated with each of the annotation instances for one of:

(i) an automatic acceptance of the at least one named entity or class annotation instance,

(ii) an automatic rejection of the at least one named entity or class annotation instance, and

(iii) the selective presentation of the at least one named entity or class annotation instance.

12. The method of claim 11, wherein:

the annotation instances above the adjusted confidence level will automatically be accepted as valid and used in a next training phase; and

the annotation instances below the adjusted confidence level will automatically be rejected as invalid.

13. The method of claim 1, wherein learning the annotator for a particular named entity or class includes using labeling schemes.

14. The method of claim 1, wherein the learned annotators are applied to text data to annotate new instances or correct previous annotations, wherein each of the at least one named entity or class annotation instance is assigned a confidence level estimating a probability that the assignment is correct.

15. The method of claim 1, wherein when the selectively presented annotations are not acceptable, the changes are made by one of:

(i) selecting specific annotation instances,

(ii) selecting an entire list of annotation instances that was presented for viewing, and

(iii) inspecting bins of the annotation instances in context, where the bins correspond to confidence level ranges.

16. The method of claim 15, wherein the bins allow a user to inspect some examples and if they are correct, choose to one of accept and reject with one action all instances in that bin.

17. The method of claim 16, wherein if the user determines some examples in a particular bin of the inspected bins are correct, all of the at least one named entity or class annotation instance can be accepted within the particular bin and all bins with higher confidence level ranges than the accepted bin such that, at one time, entire groups of all the at least one named entity or class annotation instance can be accepted.

18. The method of claim 16, wherein if the user determines some examples in a particular bin of the inspected bins are incorrect, all of the at least one named entity or class annotation instance can be rejected within the particular bin and all bins with lower confidence level ranges than the rejected bin such that, at one time, entire groups of all the at least one named entity or class annotation instance can be rejected.

19. The method of claim 1, further comprising correcting the at least one named entity or class annotation instance by deleting annotation instances, rebracketing annotation instances, relabeling annotation instances, adding or deleting annotation instances or any combination of rebracketing and relabeling.

20. The method of claim 1, wherein one of:

at each stage of learning in the iterative learning step, previously learned annotators are discarded and entirely new annotators are learned from current training data, and

at each stage of learning in the iterative learning step, previously learned annotators are updated.

21. The method of claim 1, further comprising correcting the annotation instances when a confidence level associated with the annotation instances falls within a predetermined range.

22. The method of claim 1, wherein confidence levels associated with each of the annotation instances is generated using the Generalized Winnow learning algorithm.

23. The method of claim 1, wherein the step of iteratively learning annotators includes the step of determining that a sequence of token level classifications and associated confidence levels constitutes an instance of a type of named entity or class.

24. The method of claim 23, wherein the determining step determines that a consecutive sequence of one or more tokens each of which is labeled with one or more of the types of named entity or class and each type assignment of which has an associated confidence level that equals or exceeds a required confidence level to be in a type of named entity or class is a candidate annotation instance of the type of named entity or class.

25. A method of learning annotators for use in an interactive machine learning system for processing electronic text, the method comprising the steps of:

providing examples of a type of a named entity and unannotated textual data; and

iteratively learning annotators based on at least one of the examples of a named entity and unannotated textual data, where at the end of each iteration, any annotation, generated from the learned annotators, having a confidence level within a confidence level range is presented for review and, if required, corrected based on feedback.

26. A method of learning annotators for use in an interactive machine learning system, the method comprising the steps of:

a user sequentially labeling annotation instances in a current document from a document set;

a machine learning algorithm concurrently training on the documents in the document set to learn at least one annotator for at least one named entity or class; and

assigning a confidence level to each of the annotation instances by the learned at least one annotator such that any annotation instance which has a confidence level that is equal to or above a predetermined confidence level threshold and that occurs in a current document being labeled will be presented to the user for review and possible action.

27. The method of claim 26, further comprising discarding the annotation instances determined by the machine learning system which fall below the predetermined confidence level threshold.

28. The method of claim 27, wherein each named entity or class type has a separate confidence level threshold.

29. The method of claim 26, wherein the machine learning system continuously updates its knowledge state based on flow of new annotations from the labeled documents and applies this knowledge state, as an updated annotator or annotators, to a current document being labeled to suggest a new or new annotations for the current document being worked on.

30. The method of claim 26, further comprising providing sample text with seeds for the type of named entity or class as training data.

31. The method of claim 26, wherein the review and possible correction step includes at least one of:

the user explicitly accepting the presented annotation instance;

the user explicitly rejecting the presented annotation instance;

the user rebracketing and explicitly accepting the presented annotation instance;

the user relabeling and explicitly accepting the presented annotation instance; and

the user rebracketing, relabeling and explicitly accepting the presented annotation instance.

32. The method of claim 26, further comprising accepting annotation instances which are not explicitly rejected by the user.

33. The method of claim 32, wherein the accepting of annotation instances not explicitly rejected by the user is accomplished implicitly by the user moving to a new document or explicitly by taking an acceptance action.

34. The method of claim 26, further comprising accepting annotation instances which were corrected, relabeled, rebracketed or added by the user.

35. An apparatus for learning annotators for use in an interactive machine learning system for processing electronic text, comprising:

a means for providing at least partially annotated text data or unannotated text data with seeds or seed models of instances of at least one named entity or class to be learned;

a means for iteratively learning annotators for the at least one named entity or class using a machine learning algorithm from the at least one named entity or class;

a means for applying the learned annotators to text data resulting in the annotation of at least one named entity or class annotation instance; and

a means for selectively presenting for review and correction, if determined, representations of annotation instances identified by the learned annotators.

36. The apparatus of claim 35, further comprising a component to export the final annotators for use in processing electronic text.

37. The apparatus of claim 35, further comprising a component to determine confidence levels associated with the individual annotation instances.

38. An apparatus for learning annotators for use in an interactive machine learning system for processing electronic text, comprising:

means for providing examples of a type of a named entity and unannotated textual data; and

means for iteratively learning annotators based on at least one of the examples of a named entity and unannotated textual data, where at the end of each iteration, any annotation, generated from the learned annotators, having a confidence level within a confidence level range is corrected based on feedback.

39. A computer program product comprising a computer usable medium having a computer readable program code embodied in the medium, the computer program product includes:

a first computer component to provide at least partially annotated text data or unannotated text data with seeds or seed models of instances of at least one named entity or class to be learned;

a second computer component to iteratively learn annotators for the at least one named entity or class using a machine learning algorithm from the at least one named entity or class;

a third computer component to apply the learned annotators to text data resulting in the annotation of at least one named entity or class annotation instance; and

a fourth computer program component to selectively present for review and correction, if determined, representations of annotation instances identified by the learned annotators.